

암 발생예측 모형과 유전위험점수에 관한 고찰

정금지^{1,2)}, 김소리울^{1,2)}, 윤미옥²⁾, 전티나²⁾, 지선하^{2)†}

¹⁾연세대학교 대학원 보건학과

²⁾연세대학교 보건대학원 역학건강증진학과, 국민건강증진연구소

Review on Genetic Risk Score and Cancer Prediction Models

Keum Ji Jung^{1,2)}, Soriul Kim^{1,2)}, Miwuk Yun²⁾, Christina Jeon²⁾, Sun Ha Jee^{2)†}

¹⁾Department of Public Health, Graduate School, Yonsei University

²⁾Department of Epidemiology and Health Promotion, Institute for Health Promotion, Graduate School of Public Health, Yonsei University

Abstract

Objectives: In genome-wide association studies (GWASs), single-nucleotide polymorphisms (SNPs) that have been identified as cancer-associated loci are common, but they confer only small increases in risk. The question was whether combining multiple disease-related SNPs and the modest effects within Genetic Risk Score (GRS) may be useful in identifying subgroups that are at high risk of cancer.

Methods: In this paper, we first reviewed articles that examined the predictability of GRS on cancer prediction models. Our data sources included a PubMed search of the literature published until February 2014. Secondly, we have calculated the GRS using the data example data with five SNPs related colorectal cancer (CRC) obtained from the Korean cancer prevention study II. Two approaches were used to calculate the GRS: a simple risk alleles count method (counted GRS) and a weighted method based on the genotype frequencies for each SNP and the effect sizes (allelic odds ratio or beta coefficient) from our study (weighted GRS).

Results: Of 31 studies initially identified, 16 (135,110 participants) met the inclusion criteria. Among 16 articles, 7 studies were related to prostate cancer, 6 studies to breast cancer, and 3 studies to colon cancer and lung cancer. Fifteen studies except for one study concluded that in general, a genetic score may be helpful or useful in identifying the high risk group and particularly to determining the high risk individual among patients within a “gray zone” of cancer risk. The weighted GRS with age and sex (AUC=0.9333) had higher predictability on the CRC risk than the model with GRS alone (AUC=0.816).

Conclusions: Although adding GRS improves prediction model performance, the clinical utility of these genetic risk models is limited. Nonetheless, the modelling suggests public health potential since it is possible to stratify the population into cancer risk categories, thereby informing targeted prevention and management.

Keywords: Single-nucleotide polymorphisms, Genetic risk score, Prediction model

[접수일: 2014년 03월 13일, 수정일: 2014년 05월 22일, 게재확정일: 2014년 05월 27일]

† Corresponding Author: Sun Ha Jee, PhD

Department of Epidemiology and Health Promotion, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-749, Korea. Tel:+82-2-2228-1523 E-mail: jsunha@yuhs.ac

* This work was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health & Welfare (1220180) and National Research Foundation of Korea grant funded by the Korea government (2011-0029348), Republic of Korea.

1. 서론

전체 사망원인의 1위를 차지하는 암은 고위험군을 사전에 확인하여 예방 및 관리하는 것이 필수적이다. 암 발생은 환경적인 요인과 유전적인 요인이 복합적으로 작용하는 것으로 알려져 있으며 [1], 환경위험 요인에는 흡연, 과도한 음주, 비만 등이 알려져 있다 [2-4].

또한 최근 전장유전체분석(genome wide association analysis, GWAS)을 통해 전립선암, 대장암, 폐암, 유방암 등에 대해서 관련 있는 여러 단일염기다형성(single-nucleotide polymorphism, SNP)들이 보고되고 있다 [5-8]. 전장유전체분석에서 밝혀진 여러 단일염기다형성의 위험도(risk)를 점수(score) 개념으로 바꾸어 유전성(heritability)이 약 80% 되는 정신분열성질환에 적용하여 예측력을 평가한 바 있다 [8]. 그러나 대부분 단일염기다형성과 질병발생과의 관련성 연구는 서양인을 대상으로 실시한 것이며 동양인이나 한국인을 대상으로 연구한 사례는 많지 않다. 더구나 암 발생에 있어서 환경요인과 유전요인의 상호작용에 대한 연구는 더욱 부족한 것이 현실이다.

지금까지 보고된 암 발생과 관련된 SNP의 역학적 관련성 연구를 살펴보았을 때 각 SNP의 질환 발생에 대한 관련성은 교차비(odds ratio, OR)로 볼 때 1.5미만이다. 또한 SNP의 minor allele 혹은 risky allele의 유병률이 낮으므로 이것이 기여하는 정도가 매우 작은 것이 특징이다. 따라서 최근에는 여러 개의 SNP의 영향을 통합하여 관련성을 보기 위한 방법으로서 유전위험점수(genetic risk score, GRS)의 개념이 도입되어 연구되고 있다 [8,9].

이 연구는 암 발생 예측력을 유전요인 즉, 여러 개의 단일염기다형성의 영향을 평가하는 방법 중에 하나로서 유전위험점수의 영향에 대한 고찰을 시도하였다. 또한 유전위험점수에 대한 사례자료 분석을 통해 유전위험점수와 대장암위험도 예측

과의 관련성을 보았다. 이를 통해 향후 암 발생예측모형 연구에 유전위험점수의 유용성에 대한 기초자료로서 활용이 가능할 것으로 기대한다.

2. 연구 방법

1) 문헌검색

이 연구는 2014년 2월 28일까지 PubMed에 보고된 영문문헌을 검색하였다. 검색에 사용된 키워드는 [genetic risk score, genetic risk, genetic score, gene score, gene risk] AND [cancer, cancer risk] AND [prediction]이었으며, 이에 일차적으로 만족하는 논문은 31편이었다. 이들 논문들로부터 저자, 연도, 국가, 대상자 수, 남녀구분, 사용된 단일염기다형성(SNP) 수, 유전위험점수 계산 방법, 예측모형(prediction model) 방법, 암 종류, 반복(replication) 연구 여부 등에 대한 자료를 추출하였다 [6,9,11,15-27].

2) 사례분석대상

이 연구는 문헌을 통해서 유전위험점수에 대한 방법론을 고찰한 후 실제 자료를 통해서 사례분석 결과를 제시하였다. 한국인 암 예방 연구 II (Korean cancer prevention study II)는 2005년 서울시 대사증후군 연구사업단에서 개발한 일반인구 코호트로서 혈액을 확보한 대상자는 153,856명이다. 153,856명 중에서 일반 인구 1,004명에 대해서 2007년 아디포넥틴 전장유전체분석을 실시하였으며, 이후 대장암 환자 325명에 대해서 추가적인 전장유전체 분석을 실시하였다 [9]. 이 연구에서는 대장암의 GWAS 분석결과에서 자료이용이 가능한 1,262명을 사례분석 대상으로 선정하였다. 사례분석용 1,262명(남자 765명, 여자 497명)중 대장암은 303명과, 대조군은 959명이 포함되었다. 이 중에서 5개 SNP 검사에 결측이 모두 없는 대상자는 1,040명(남자 617명, 여자 423명)중 대장암 184명,

대조군 856명이었다. 한국인 암 예방 연구 II는 세브란스병원 임상연구심의(4-2011-0277)를 통과하였으며, 연구에 사용된 모든 대상자로부터 서면 동의와 유전자 검사동의서를 받았다.

3) 유전자 분석

대장암 전장유전체분석은 주)DNALink에서 Affymetrix Genomewide Human SNP Array 5.0 (Affymetrix Inc., Santa Clara, CA, USA)를 사용하여 검사하였다. 내부정도관리목적으로 call rate는 95%이상을 요구하였다. 자세한 유전자분석에 대해서는 이전 연구에서 소개하였다 [9]. 대장암 전장유전체 분석에서 확인된 상위 5개 SNP는 rs12613322, rs16846499, rs17140038, rs6962549, rs824575 이다.

4) 분석방법

자료 분석은 두 가지 방법으로 시행하였다. 첫째, 문헌검색을 통하여 유전위험점수와 암 위험도관련 연구의 특성을 고찰하였다. 문헌에서 얻어진 개별 논문에 대해서 발표연도, 국가, 대상자 수, 남녀 대상자 비율, 평균연령, 예측모형에 사용된 SNP 수, 사용된 통계모형, 질환 명, 예측모형의 replication 여부 등에 대한 특성을 분석하였다. 둘째, 유전위험점수의 사례분석을 위하여 유전위험점수를 계산하였다. 또한 5개 SNP를 사용하여 계산한 유전위험점수와 대장암 위험도 예측은 다음과 같은 방법으로 시행하였다. 우선 유전위험점수만을 포함한 예측한 모형과 유전위험점수에 연령과 성을 포함한 모형을 각각 비교하였다.

이 연구에서 모형의 개발과 평가를 서로 다른 자료를 사용하였다. 즉, 모형의 개발은 전체자료 1,040명중에서 80%만 무작위 추출한 832명이었고, 모형의 평가는 나머지 20%인 208명으로 실시하였다. 모형의 예측력 평가는 Areas under the receiver operating characteristic (ROC)

curves (AUC)로 시행하였다. 통계분석은 SAS 9.2 (SAS Institute Inc., Cary, NC, USA)를 사용하였다.

3. 연구 결과

1) 유전위험점수(GRS) 관련 연구문헌의 일반적 특성

일차적으로 확인된 31개 논문 중에서 유전위험점수와 암 위험도와 관련성을 보고한 논문은 16편이었다. Table 1은 연구에 포함된 16개 문헌의 특성을 요약한 것이다. 연도별로 2011년에 1편, 2012년에 8편, 2013년에 7편의 논문이 발표되었다. 16개 연구의 전체 대상자 수는 135,110명이었고, 평균 8,444명(범위 308명-44,389명)이었다. 암 종별로는 전립선암 연구가 7편으로 가장 많았고, 유방암 연구가 5편이었고, 나머지는 대장암과 폐암에 관한 연구였다.

Table 2는 16개 문헌으로부터 유전위험점수 개발에 사용한 SNP의 수, 방법, 그리고 유전위험점수와 암 위험과의 관련성을 요약하였다. 연구에서 사용한 평균 SNP수는 19개(범위 5개-38개)였다. 유전위험점수는 counted 방법을 사용한 연구가 6편, weighted 방법 5편, 두 가지 모두를 사용한 연구 4편, 그리고 1편은 방법을 구체화하지 않았다.

유전위험점수의 유용성에 대해서는 16개 연구 중 1편(Tindall et al. [24])을 제외하고 15편 연구에서 암을 분류하거나 예측하는데 유용하거나 도움이 된다고 결론지었다. 그러나 대부분 연구에서 유전위험점수가 유용성을 보이긴 했지만 관련성이나 예측력이 약했고 임상적 유용성은 제한적이라고 기술하였다.

2) 유전위험점수(GRS) 개발 방법

(1) Counted GRS

Counted GRS 계산법은 SNP마다 단순히 risky allele의 수를 더하는 방법이다. 예를 들어, SNP의 allele이 A와 B가 있다면, 유전자 다형은 AA, AB,

BB가 된다. 이때, B가 risky allele 이라면 B의 수 만큼 1을 할당하면 0, 1, 2개가 된다 (수식 1). 일반적으로 minor allele이 risky allele과 동일하지만 그렇지 않은 경우도 있다.

수식 1

$$\text{Counted GRS} = \sum_{i=1}^n \text{Number of risky allele in } SNP_i$$

n : 전체 대상자수, SNP_i : 각각의 SNP

(2) Weighted GRS

Weighted GRS 계산은 SNP의 유전자형에 따라 가중치를 준 후 합산하는 방법이다. 가장 많이 사용하는 방법은 SNP와 질환과의 관련성에 대한 메타분석이나 대규모 역학연구(권소시업 등)로 부터 얻어진 회귀계수 값을 사용하여 가중치를 주는 방법이다. 예를 들어, SNP의 allele이 A와 B가 있고, B가 risky allele이라면 유전자다형 AA, AB, BB에 대해서 risky allele 수는 각각 0, 1, 2가 될 것이다. 그런 다음, additive 모형을 가정하여 allele 증가 당 회귀계수 값을 구한 후 risky allele 수, 즉 0, 1, 2에 가중치로서 회귀계수를 곱한 후 더하는 방법이다 (수식 2).

수식 2

$$\text{Weighted GRS} = \sum_{i=1}^n \text{Number of risky allele in } SNP_i \times \text{Weight}_i$$

n : 전체 대상자수, Weight_i : i 번째 대상자의 가중치, SNP_i : 각각의 SNP

3) 사례분석결과

이 연구는 앞서 정리한 두 가지 유전위험점수 계산 방법을 한국인 암 예방 연구 II의 실제 자료에 적용 후 사례분석을 시도해 보았다. Table 3은 대장암 관련 5개 SNP를 사용하여 additive 가정 하에 유전위험점수를 계산하고, 이를 대장암 위험도 예측에 적용한 결과이다. 대장암 관련 5개 SNP는 대장암 325명과 대조군 1,004명에 대한 GWAS 분석

결과에서 상위 5개에 속하는 것이다 [9]. 이 연구의 사례분석에는 SNP별 결측치가 다르므로, 대장암과 대조군의 수의 차이를 보였다 (Table 3). 분석에 사용된 5개 SNP는 대장암에서 상위 5개에 해당되는 매우 유의한 SNP이므로 대장암과의 관련성(OR)은 매우 강하였다. Table 4는 각 유전자형의 영향에 차이가 있을 것으로 예상하여 codominant 혹은 dominant 가정 하에 유전위험점수를 계산하기 위한 관련성 분석한 것이다. 즉, 5개의 SNP중 1개만 제외하고 모두 dominant 형을 보이는 것이 특징이었다.

사례분석에서 얻어진 risky allele의 수와 회귀계수 β 값을 가지고 counted GRS와 weighted GRS를 각각 계산하였다. 즉, counted GRS은 개인별로 각 SNP의 risky allele의 수를 더하여 계산하였다. Weighted GRS는 각 SNP의 risky allele 수에 β 값을 곱한 후 더하여 계산하였다.

Figure 1은 5개 SNP를 가지고 사례분석 자료에 적용하여 계산한 유전위험점수의 분포를 보여주는 것이다. 분석에 사용된 수는 5개 SNP 결과에 결측이 없었던 1,040명이었다. Figure 1의 왼쪽은 counted GRS를 대장암 환자군과 대조군에 따라 그린 것이고, 오른쪽 그림은 weighted GRS를 대장암 환자군과 대조군에 따라 그린 것이다. 양쪽 그림 모두 대조군의 경우보다 대장암 환자의 경우에 GRS 값의 분포가 오른쪽으로 치우쳐 있는 것을 볼 수 있었다.

Figure 2는 앞서 계산한 counted GRS와 weighted GRS를 사용하여 대장암의 예측력을 AUC 값으로 표현하고 있다. Figure 2에서 연령과 성만으로 예측한 모형(AUC, 0.831)을 양쪽 그래프에 제시하고 있다. Figure 2의 왼쪽은 counted GRS만으로 예측한 모형(AUC, 0.807)과 counted GRS에다 연령과 성을 포함한 모형으로 예측한 모형(AUC, 0.931)을 각각 제시하고 있다. Figure 2의 오른쪽은 weighted GRS에 대한 결과로서 앞서 소개한 두 모형에 대한 AUC값은 각각 0.816, 0.933이었다.

즉, counted GRS와 weighted GRS 모형의 예측력에는 차이가 없었으며, 연령과 성만으로 된 예측력을 제외하면 GRS자체만으로 약 10% 추가적인 예측력을 보였다.

GRS에 연령과 성을 포함한 대장암 위험도 모형은 로지스틱 모형이었다. 즉, counted GRS 모형의 대장암 위험도는 CP1 (수식 3), weighted GRS 모형의 대장암 위험도는 WP1 (수식 4)로 각각 제시하였다.

수식 3

$$CP = -6.6125 + 0.1126 \times AGE - 0.7962 \times SEX + 1.9821 \times CGRS$$

$$CP1 = \frac{e^{CP}}{1 + e^{CP}}$$

여기서, CP1은 대장암의 위험도 점수, age는 연

령, sex는 성(남자 0, 여자 1), CGRS는 counted GRS를 각각 의미한다.

수식 4

$$WP = -6.7111 + 0.1137 \times AGE - 0.7859 \times SEX + 0.8034 \times WGRS$$

$$WP1 = \frac{e^{WP}}{1 + e^{WP}}$$

여기서, WP1은 대장암의 위험도 점수, age는 연령, sex는 성(남자 0, 여자 1), WGRS는 weighted GRS를 각각 의미한다.

사례분석 결과를 종합하면, counted GRS 모형과 weighted GRS 모형이 대장암 예측에 주는 영향은 큰 차이가 없었다. GRS에 연령과 성을 포함할 경우에도 마찬가지로 대장암의 예측력이 증가하는 양상을 보였다.

Table 1. Characteristics of studies included in literature review of cancer prediction models

	Author	Year	Country	Cancer	Prediction	Replication	Study participants			
							Total subject	% of male	Case	Control
1	McCarthy AM et al.	2013	USA	Breast	Y	N	810	0.0	810	0
2	Pearce CL et al.	2013	USA	Ovarian	Y	N	12,940	0.0	5,566	7,374
3	Rinella ES et al.	2013	USA	Breast	Y	Y	533	0.0	239	294
4	Jiang H et al.	2013	China	Prostate	Y	N	308	100.0	141	167
5	Purrington KS et al.	2013	Consortium study	Breast	Y	N	4,928	0.0	1,529	3,399
6	Tindall EA et al.	2013	South Africa	Prostate	Y	N	837	100.0	522	315
7	Butoescu V et al.	2013	Netherlands	Prostate	Y	N	346	100.0	191	155
8	He C et al.	2012	Consortium study	Breast	Y	N	37,857	0.0	3,683	34,174
9	Darabi H et al.	2012	Sweden	Breast	Y	N	1,890	0.0	1,022	868
10	Li H et al.	2012	China	Lung	Y	N	5,068	26.6	2,283	2,785
11	Zheng J et al.	2012	China	Prostate	Y	N	2,633	100.0	1,108	1,525
12	Dunlop MG et al.	2012	Consortium study	Colon	Y	N	44,389	NA	24,395	19,994
13	Ribeiro RJT et al.	2012	Portugal	Prostate	Y	N	1,006	100.0	449	557
14	Lindström S et al.	2012	Consortium study	Prostate	Y	N	15,161	100.0	7,509	7,652
15	Jo J et al.	2012	Korea	Colon	Y	N	1,163	59.1	187	976
16	Aly M et al.	2011	Sweden	Prostate	Y	N	5,241	100.0	2,135	3,106

NA: not available, Y: yes, N: no

Table 2. Main results from studies in literature review of cancer prediction models

	Author	Year	SNP	Method of GRS	Summary of incremental impact of GRS on risk of prediction model
1	McCarthy AM et al.	2013	12	Weighted	The addition of the GRS to BCRAT (Gail model) reclassifies the high risk status of some women undergoing mammography.
2	Pearce CL et al.	2013	6	Weighted	No interaction between the GRS and the environmental risk factors on ovarian cancer.
3	Rinella ES et al.	2013	7	Counted	The GRS may improve breast cancer risk.
4	Jiang H et al.	2013	24	Weighted	The GRS may be helpful to determine the need for prostate biopsy among patients within a 'gray zone' of PCa risk.
5	Purrrington KS et al.	2013	30	Weighted	GRS had better discriminatory accuracy (AUC = 0.64) compared with that of the Gail model (0.50) applied in the Women's Health Initiative.
6	Tindall EA et al.	2013	38	Both	Genetic risk factors show no evidence for PCa prediction
7	Butoescu V et al.	2013	9	Counted	The predictive performance of the clinical model was only slightly improved by adding GRS.
8	He C et al.	2012	19	Both	The GRS was associated with breast cancer risk and may contribute modestly to breast cancer risk prediction.
9	Darabi H et al.	2012	18	Counted	The GRS had moderate improvements.
10	Li H et al.	2012	5	Both	The GRS related to lung cancer added moderate accuracy.
11	Zheng J et al.	2012	33	Weighted	The GRS related to PCa risk may be useful for risk prediction.
12	Dunlop MG et al.	2012	10	Counted	The GRS provided additional information on CRC risk prediction.
13	Ribeiro RJT et al.	2012	29	Not specified	The addition of the GRS to PSA improved the predictive accuracy.
14	Lindström S et al.	2012	25	Counted	The GRS can be useful for identifying younger men that might benefit from PSA screening.
15	Jo J et al.	2012	8	Both	GRS improve the prediction of CRC when considered in conjunction with age and family history.
16	Aly M et al.	2011	35	Counted	Although the genetic model reduced the number of biopsies, the clinical significance of this finding requires further evaluation.

SNP: single-nucleotide polymorphism, GRS: genetic risk score, PCa: prostate cancer, CRC: colorectal cancer, PSA: prostate specific antigen

Table 3. Allelic odds ratios on colorectal cancer in the Korean cancer prevention study II*-additive model

SNPs	Chr.	Gene	Reference allele	Risky allele	Frequency of risky allele	Case	Control	Age and sex adjusted*			
								β	OR	95% CI	p-value
rs12613322	2	Intergenic	A	C	0.053	209	751	1.99	7.28	4.26 - 12.45	<0.001
rs16846499	2	-	T	A	0.032	202	769	2.70	14.86	7.31 - 30.21	<0.001
rs17140038	5	SEMA6A	A	C	0.051	206	757	2.48	11.99	6.51 - 22.09	<0.001
rs6962549	7	-	G	A	0.028	194	771	2.58	13.14	6.53 - 26.44	<0.001
rs824575	16	Intergenic	C	T	0.042	226	741	2.55	12.85	6.98 - 23.66	<0.001

* 1,006 subjects (80% random sample of total subjects (n=1,262) were used to develop the prediction model.

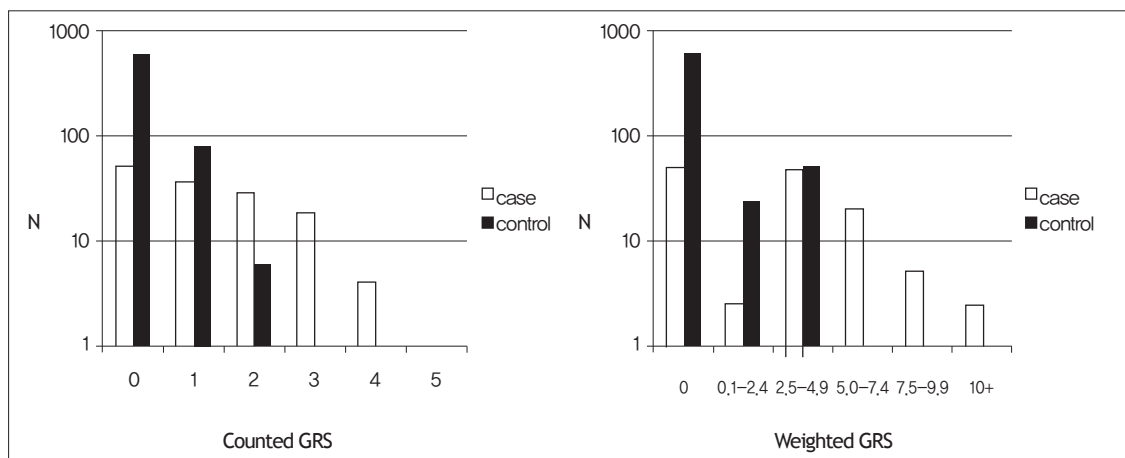
SNPs: single-nucleotide polymorphisms, Chr.: chromosome, OR: odds ratio, CI: confidence interval

Table 4. Odds ratios between genotype and colorectal cancer in the Korean cancer prevention study II*

SNPs	Reference allele	Risky allele	Case	Control	Model of inheritance	Age and sex adjusted*						
						Reference	Heterozygote			Homozygote		
							β	OR	95% CI	β	OR	95% CI
rs12613322	A	C	209	751	Codominant	AA	2.11	8.24	4.74-14.31	0.91	2.50	0.09-71.74
					C dominant		2.11	8.22	4.73-14.28			
rs16846499	T	A	202	769	Codominant	TT	2.69	14.73	7.21-30.09	14.69	>999	NE
					A dominant		2.68	14.54	7.12-29.69			
rs17140038	A	C	206	757	Codominant	AA	2.64	14.00	7.28-26.92	3.31	27.37	2.79-268.36
rs6962549	G	A	194	771	Codominant	GG	3.07	21.49	9.91-46.63	1.90	6.71	0.83-54.20
					A dominant		3.05	21.14	9.75-45.81			
rs824575	C	T	226	741	Codominant	CC	2.72	15.23	8.05-28.80	2.53	12.52	1.03-152.89
					T dominant		2.70	14.95	7.91-28.28			

* 832 subjects (80% random sample of total subjects (n=1,040) were used to develop the prediction model.

SNPs: single-nucleotide polymorphisms, OR: odds ratio, CI: confidence interval

**Figure 1.** Distribution of counted and weighted genetic risk score (GRS) in the Korean cancer prevention study II.

*832 subjects (80% random sample of total subjects (n=1,040) were used to develop the prediction model.

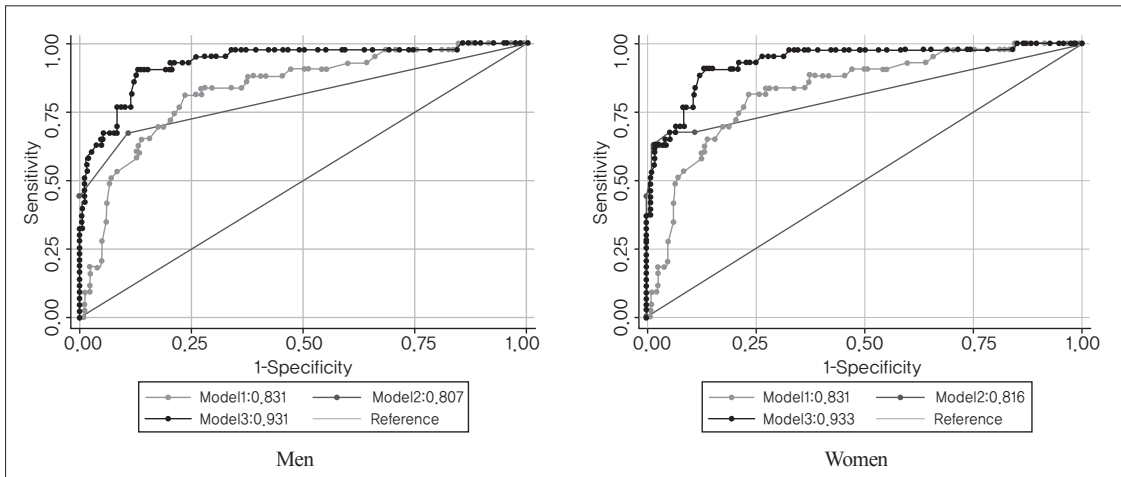


Figure 2. Areas under the receiver operating characteristic (ROC) curves of genetic risk score on colorectal cancer in the Korean cancer prevention study II.

*208 subjects (20% random sample of total subjects (n=1,040) were used to test the prediction model.

Model 1: Risk model with age and sex, Model 2: Risk model with counted GRS alone, Model 3: Risk model with age, sex and counted GRS

4. 고 찰

이 연구는 유전위험점수(GRS)가 암 발생 예측 모형에 주는 유용성을 평가하고자 관련 문헌연구를 시도하였다. 또한 GRS의 사례분석을 위해서 한국인 암 예방연구 II 자료를 분석하였다. 이 연구에서 확인한 16편 연구 중 15편 연구에서 GRS는 암 발생과의 분명한 관련성을 보였다. 그러나 관련성의 크기가 비교적 작고, 유전요인의 유병률도 낮아서 예측에 기여하는 분율이 작았다. 그렇지만 장기적으로 GRS는 중앙 바이오마커나 임상에서 사용하는 인자들과 더해졌을 때, 예측력을 증가시킬 수 있으며, 특히, 정상과 비정상의 중간에 놓여있는 회색지대(gray zone)의 비특이성을 개선하는데 도움을 줄 수 있는 가능성이 있다고 결론지었다.

또한 한국인 암 예방 연구 자료 1,040명을 이용하여 실제로 대장암에 대한 GRS를 계산하고 관련성을 본 결과 counted GRS와 weighted GRS에 있

어서 예측력의 차이는 없었다. 오히려, GRS에 연령과 성을 포함한 모형으로 확장하였을 때 대장암의 예측력(AUC)은 약 0.9333까지 증가하였으며, 연령과 성에 GRS가 추가됨으로서 약 10%의 예측력이 증가되었다 (Figure 2).

1) 유전위험점수를 이용한 암 예측모형에 관한 연구 문헌의 요약

이 연구를 통해 확인한 16편의 관련 연구문헌을 요약하면 다음과 같았다. 첫째, 유전위험점수가 발연구에서 사용된 SNP는 주로 이전 GWAS연구에서 보고된 경우이거나 후보 유전자에서 보고된 SNP를 선정하였다. 또한, GWAS 연구 혹은 별도의 replication 연구에서 얻어진 OR나 회귀계수가 GRS계산에 사용되는 것이 일반적이었다. 둘째, GRS를 이용한 예측모형에 환경적인 요인을 포함한 연구는 거의 없었다. 암 발생은 유전요인과 환경적인 요인이 함께 상호 작용하여 발생하는 것은 잘 알려진 사실이다. 그렇지만 암 발생 예측 모형

을 개발하는 데 있어서 암 발생 위험요인으로 잘 알려진 흡연, 음주, 운동, 식이 습관 등을 포함한 모형은 매우 드물다. 한편의 연구에서 GRS 연구에 음주, 흡연, 운동여부 등을 포함한 후 simulation 연구를 한 결과 두 개 이상의 환경위험요인이 추가될 경우 예측력을 증가될 수 있었음을 시사한 바 있었다 [10]. 셋째, GRS는 counted GRS와 weighted GRS 두 가지 방법이 있으며, 일부 문헌에서 counted GRS보다 weighted GRS가 예측력이 우수하다고 보고하였다 [11]. 즉, Li et al. [11]은 폐암 예측 연구에서 5개 SNP를 이용하여 중국인 5,068명의 환자-대조군 연구를 수행한 결과 weighted GRS가 counted GRS 보다 폐암을 더 잘 예측하였다. 그러나 본 연구에서 실제자료를 통한 사례분석결과 두 방법에 차이를 볼 수 없었다. 넷째, 개발된 GRS를 포함한 예측모형의 replication 연구가 부족한 상황이다. 대부분의 연구가 자체 환자-대조군 연구로 예측력을 제시하였으나, 이 모형을 또 다른 자료에 적용하여 예측력을 제시한 replication 연구는 없었다. 암 발생 예측 모형의 일반화를 위해서는 서로 다른 연구대상에 적용하여 반복 평가 연구를 실시하는 것이 필수적이다.

2) 기존 예측모형에 유전위험점수(GRS)를 추가한 연구 사례

유방암 위험도 예측모형은 이미 백인에서 개발되어 사용되고 있는 Gail 모형이 있으며 [12], 이를 수정한 것이 BCRAT라는 모형이다 [13]. BCRAT(breast cancer risk assessment tool)은 Gail 모형으로 알려진 것과 동일하다 [14]. McCarthy et al. [15]은 40세 이상 백인과 흑인에서 맘모그래피를 수행한 여성을 대상으로 기존 유방암 예측모형에다 SNP를 포함한 모형을 평가하였다. 즉, BCRAT라는 유방암 예측모형에다 12개의 SNP를 추가하기 전과 후의 5년 발생 위험도 결과를 평가하였다. 이 연구는 유방암 고위험 군을 평가하기 위한 것으로서 통계적

평가는 kappa를 사용하였다. 최종 분석에 포함된 대상자는 810명 여성(흑인 39%, 백인 55%)이었고 BCRAT로부터 5년 평균 유방암 위험은 흑인 1.18%, 백인 1.71%였다. 결론적으로 맘모그래피를 받은 여성에게 12개 SNP로 구성된 유전위험점수를 포함한 모형을 적용할 경우 고위험군 발견이 증대되었다. 비슷한 시도로서 Darabi et al. [16]은 18개 SNP를 이용하여 스웨덴에서 수정 보완한 Gail모형에 적용한 결과 유방암 환자 예측(AUC) 값을 GRS 포함 전 0.548 (0.527-0.568)에서 GRS 포함 이후에 0.615 (0.595-0.634)로 증가하였다.

Rinella et al. [17]은 BRCA 1/2 mutation은 없으면서 가족력을 가진 고위험군을 대상으로 GRS의 예측력을 평가하였다. 저자들은 유방암의 조기발견이나 치료에 유전위험 모형은 중요하지만, 이러한 모형이 임상적 유용성을 얻으려면 이전에 확인된 GWAS를 넘어서 새로운 유전인자가 밝혀져야 한다고 주장하였다. 연구팀은 새롭게 선정한 19개 SNP를 통해서 GRS를 개발하고, 이를 통해 유방암 예측력을 평가한 결과 예측력 AUC는 0.74 (0.69-0.79)로 이전 연구에서도 높았다.

3) 기존 바이오 마커에 유전위험점수(GRS)를 추가한 연구 사례

전립선 특이 항원(prostate specific antigen, PSA)은 전립선암을 예측하는 바이오 마커로 잘 알려져 있다. 그러나 PSA는 전립선암을 예측하고 진단하는 데 있어서 비 특이적이라는 논란이 있어 왔다. 연구자들은 이러한 PSA의 제한점을 보완하기 위하여 전립선암과 관련 GWAS 연구나 후보유전자로부터 얻은 SNP의 도움을 받고자 연구를 시도하였다. Ribeiro et al. [18]은 전립선암과 관련된 29개 SNP를 선정하고, 연령과 PSA를 포함한 모형에 유전위험점수를 더한 결과 예측력이 증가하였다고 보고하였다. Lindström et al. [19]

은 7,509명 전립선암 환자와 7,652명 대조군을 대상으로 25개 SNP를 통해서 GRS의 예측력을 평가한 결과 임상적인 유용성은 제한적이라고 보고하였다. Aly et al. [20]은 전립선암 관련 35개 SNP를 통해서 연구한 결과 Lindström et al. [19]의 연구와 비슷한 결론을 내렸다.

전립선암은 최근에 매우 증가하고 있는 암 가운데 하나이고, 과도한 생검이 새로운 보건문제로 대두되는 상황에서, 이를 해결할 방법으로 GRS를 적용한 연구가 있다. Jiang et al. [6]은 전립선암 진단을 위한 생검을 받은 중국인 308명 대상으로 연구를 수행하였다. 즉, 생검을 받은 308명 중 141명 (45.8%)이 전립선암 환자로 진단되었다. 또한 연구팀이 선정한 24개 SNP가 전립선암 생검의 결과를 예측하는지를 평가하는지와 현존하는 임상변수의 예측력을 증대하는지를 평가하였다. 평균 GRS는 전립선암에서 1.30, 정상인에서 0.89로 유의한 차이가 있었다. 연령과 PSA 값을 보정한 후에 GRS가 높을수록 생검에서 전립선암 예측률이 증가하였다 ($p=0.007$). 특히 PSA 값이 1.6–20 ng/mL 정도인 대상자에서 유전위험점수는 전립선 암 예측에 기여하였고, 일단 PSA 값이 20 ng/mL 이상이면 예측에 도움이 되지 않았다.

대장암의 경우 기존 바이오마커에 GRS를 포함하여 연구한 보고는 찾을 수 없었다. 단, GRS에다 연령, 성, 가족력을 추가하여 대장암의 예측력을 평가한 결과, GRS는 예측력 증대에 도움이 안 되었다고 보고한 바 있다 [21].

이러한 연구가 계속 발전할 경우 전립선암 진단을 위한 생검을 시행한 환자에서 회색지대에 속한 대상자들을 좀 더 정확하게 진단하는데 도움을 줄 수 있을 것으로 기대한다. 아직 국내에서는 GRS와 암 발생 예측력과의 관계를 연구한 사례는 매우 드문 상태이다.

이 연구의 제한점은 일반적으로 문헌연구를 위해 검색원을 선정할 때 미국국립의학도서관 (National Library of Medicine, NLM)이 제시한

COSI (COre, Standard, Ideal) 모델을 사용하고 있다. 이중 COre는 문헌검색의 핵심이 되는 부분으로 관련 국내문헌, 핵심 데이터베이스(예를 들어 PubMed, EMBASE, Cochrane central 등)를 포괄적으로 상세히 검색하여야 한다. 그러나 이 연구의 목적은 GRS와 특정 질환과의 관련성에 대한 메타분석이 아닌, 방법론적인 고찰이었다. 따라서 PubMed에 발표된 문헌만을 고찰하였으며, 이는 문헌 검색자료 수집에 선택적 오차를 범할 수 있으므로 연구결과 해석에 주의할 필요가 있다고 본다. 또한 GRS에 사용된 5개 SNP는 대장암 GWAS 분석결과 상위 5개에 해당되는 것으로서 통계적으로 매우 의미 있는 SNP이었지만, 타당도 검증이 되지 않은 것이었다. 이는 GRS 계산의 사례분석에 사용은 가능하지만, 이 결과를 일반화할 수 없는 것임을 밝혀둔다. 더불어 GRS 연구를 위한 SNP 선정은 처음부터 자료셋이 분리되어 선정되어야 하지만, 이 연구는 같은 데이터셋을 사용함으로써 GRS의 결과가 실제보다 과도하게 추계되었을 가능성이 높다.

앞으로 국내에서도 타당도가 입증된 SNP를 통해서 GRS와 질병발생위험도 연구가 수행되어야 할 것으로 생각한다.

5. 결 론

최근 GWAS 연구를 통해 암 발생과 관련 있는 유전자 정보가 많이 발표되고 있다. 이러한 유전 정보가 기존에 임상이나 조기진단용으로 사용 중인 종양 바이오마커나 암 발생위험 예측모형이 가지고 있는 제한점을 줄이는데 유용한지 알아보는 연구가 수행되어야 할 것이다. 또한 임상적 유용성이 입증되더라도 유전자 비용을 감안한 비용-효과분석에 관한 경제성 평가 연구도 필요할 것으로 판단된다.

References

- [1] Chu YH, Hsieh MJ, Chiou HL, Liou YS, Yang CC, Yang SF, Kuo WH. MicroRNA gene polymorphisms and environmental factors increase patient susceptibility to hepatocellular carcinoma. *PloS One* 2014; 26:9(2): e89930.
- [2] Cerhan JR, Moore SC, Jacobs EJ, Kitahara CM, Rosenberg PS, Adami HO, Ebbert JO, English DR, Gapstur SM, Giles GG, Horn-Ross PL, Park Y, Patel AV, Robien K, Weiderpass E, Willett WC, Wolk A, Zeleniuch-Jacquotte A, Hartge P, Bernstein L, Berrington de Gonzalez A. A pooled analysis of waist circumference and mortality in 650,000 adults. *Mayo Clinic Proceedings* 2014; 89(3): 335-345.
- [3] Morrison DS, Parr CL, Lam TH, Ueshima H, Kim HC, Jee SH, Murakami Y, Giles G, Fang X, Barzi F, Batty GD, Huxley RR, Woodward M. Behavioural and metabolic risk factors for mortality from colon and rectum cancer: analysis of data from the Asia-Pacific Cohort Studies Collaboration. *Asian Pacific Journal Cancer of Prevention* 2013; 14(2): 1083-1087.
- [4] Blakely T, Barendregt JJ, Foster RH, Hill S, Atkinson J, Sarfati D, Edwards R. The association of active smoking with multiple cancers: national census-cancer registry cohorts with quantitative bias analysis. *Cancer Causes & Control* 2013; 24(6): 1243-1255.
- [5] Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, Hoening M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PD, Stratton MR, Dunning AM, Rahman N, Easton DF. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nature Genetics* 2010; 42(6): 504-507.
- [6] Jiang H, Liu F, Wang Z, Na R, Zhang L, Wu Y, Zheng J, Lin X, Jiang D, Sun J, Zheng SL, Ding Q, Xu J. Prediction of prostate cancer from prostate biopsy in Chinese men using a genetic score derived from 24 prostate cancer risk-associated SNPs. *The Prostate* 2013; 73(15): 1651-1659.
- [7] Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, Dong Q, Zhang Q, Gu X, Vijayakrishnan J, Sullivan K, Matakidou A, Wang Y, Mills G, Doheny K, Tsai YY, Chen WV, Shete S, Spitz MR, Houlston RS. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nature Genetics* 2008; 40(5): 616-622.
- [8] The International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; 460(7256): 748-752.
- [9] Jo J, Nam CM, Sull JW, Yun JE, Kim SY, Lee SJ, Kim YN, Park EJ, Kimm H, Jee SH. Prediction of colorectal cancer risk using a genetic risk score: the Korean Cancer Prevention Study-II (KCPS-II). *Genomics & Informatics* 2012; 10(3): 175-183.
- [10] Yarnall JM, Crouch DJ, Lewis CM. Incorporating non-genetic risk factors and behavioural modifications into risk prediction models for colorectal cancer. *Cancer Epidemiology* 2013; 37(3): 324-329.
- [11] Li H, Yang L, Zhao X, Wang J, Qian J, Chen H, Fan W, Liu H, Jin L, Wang W, Lu D. Prediction

- of lung cancer risk in a Chinese population using a multifactorial genetic model. *BMC Medical Genetics* 2012; 10(13): 118.
- [12] Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, Mulvihill JJ. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *Journal of the National Cancer Institute* 1989; 81(24): 1879-1886.
- [13] Gail MH. Discriminatory accuracy from single-nucleotide polymorphisms in models to predict breast cancer risk. *Journal of the National Cancer Institute* 2008; 100(14): 1037-1041.
- [14] Mealiffe ME, Stokowski RP, Rhees BK, Prentice RL, Pettinger M, Hinds DA. Assessment of clinical validity of a breast cancer risk model combining genetic and clinical information. *Journal of the National Cancer Institute* 2010; 102(21): 1618-1627.
- [15] McCarthy AM, Armstrong K, Handorf E, Boghossian L, Jones M, Chen J, Demeter MB, McGuire E, Conant EF, Domchek SM. Incremental impact of breast cancer SNP panel on risk classification in a screening population of white and African American women. *Breast Cancer and Research Treatment* 2013; 138(3): 889-898.
- [16] Darabi H, Czene K, Zhao W, Liu J, Hall P, Humphreys K. Breast cancer risk prediction and individualised screening based on common genetic variation and breast density measurement. *Breast Cancer Research* 2012; 14(1): R25.
- [17] Rinella ES, Shao Y, Yackowski L, Pramanik S, Oratz R, Schnabel F, Guha S, LeDuc C, Campbell CL, Klugman SD, Terry MB, Senie RT, Andrulis IL, Daly M, John EM, Roses D, Chung WK, Ostrer H. Genetic variants associated with breast cancer risk for Ashkenazi Jewish women with strong family histories but no identifiable BRCA1/2 mutation. *Human Genetics* 2013; 132(5): 523-536.
- [18] Ribeiro RJ, Monteiro CP, Azevedo AS, Cunha VF, Ramanakumar AV, Fraga AM, Pina FM, Lopes CM, Medeiros RM, Franco EL. Performance of an adipokine pathway-based multilocus genetic risk score for prostate cancer risk prediction. *PLoS One* 2012; 7(6): e39236.
- [19] Lindström S, Schumacher FR, Cox D, Travis RC, Albanes D, Allen NE, Andriole G, Berndt SI, Boeing H, Bueno-de-Mesquita HB, Crawford ED, Diver WR, Gaziano JM, Giles GG, Giovannucci E, Gonzalez CA, Henderson B, Hunter DJ, Johansson M, Kolonel LN, Ma J, Le Marchand L, Pala V, Stampfer M, Stram DO, Thun MJ, Tjonneland A, Trichopoulos D, Virtamo J, Weinstein SJ, Willett WC, Yeager M, Hayes RB, Severi G, Haiman CA, Chanock SJ, Kraft P. Common genetic variants in prostate cancer risk prediction-results from the NCI Breast and Prostate Cancer Cohort Consortium (BPC3). *Cancer Epidemiology, Biomarkers & Prevention* 2012; 21(3): 437-444.
- [20] Aly M, Wiklund F, Xu J, Isaacs WB, Eklund M, D'Amato M, Adolfsson J, Grönberg H. Polygenic risk score improves prostate cancer risk prediction: results from the Stockholm-1 cohort study. *European Urology* 2011; 60(1): 21-28.
- [21] Dunlop MG, Tenesa A, Farrington SM, Ballereau S, Brewster DH, Koessler T, Pharoah P, Schafmayer C, Hampe J, Völzke

- H, Chang-Claude J, Hoffmeister M, Brenner H, von Holst S, Picelli S, Lindblom A, Jenkins MA, Hopper JL, Casey G, Duggan D, Newcomb PA, Abuli A, Bessa X, Ruiz-Ponte C, Castellvi-Bel S, Niittymäki I, Tuupanen S, Karhu A, Aaltonen L, Zanke B, Hudson T, Gallinger S, Barclay E, Martin L, Gorman M, Carvajal-Carmona L, Walther A, Kerr D, Lubbe S, Broderick P, Chandler I, Pittman A, Penegar S, Campbell H, Tomlinson I, Houlston RS. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42,103 individuals. *Gut* 2013; 62(6): 871-881.
- [22] Pearce CL, Rossing MA, Lee AW, Ness RB, Webb PM; for Australian Cancer Study (Ovarian Cancer); Australian Ovarian Cancer Study Group, Chenevix-Trench G, Jordan SM, Stram DA, Chang-Claude J, Hein R, Nickels S, Lurie G, Thompson PJ, Carney ME, Goodman MT, Moysich K, Hogdall E, Jensen A, Goode EL, Fridley BL, Cunningham JM, Vierkant RA, Weber RP, Ziogas A, Anton-Culver H, Gayther SA, Gentry-Maharaj A, Menon U, Ramus SJ, Brinton L, Wentzensen N, Lissowska J, Garcia-Closas M, Massuger LF, Kiemeny LA, Van Altena AM, Aben KK, Berchuck A, Doherty JA, Iversen E, McGuire V, Moorman PG, Pharoah P, Pike MC, Risch H, Sieh W, Stram DO, Terry KL, Whittemore A, Wu AH, Schildkraut JM, Kjaer SK; Ovarian Cancer Association Consortium. Combined and interactive effects of environmental and GWAS-identified risk factors in ovarian cancer. *Cancer Epidemiology, Biomarkers & Prevention* 2013; 22(5): 880-890.
- [23] Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, Carpenter J, Chang-Claude J, Martin NG, Montgomery GW, Kristensen V, Anton-Culver H, Goodfellow P, Tapper WJ, Rafiq S, Gerty SM, Durcan L, Konstantopoulou I, Fostira F, Vratimos A, Apostolou P, Konstanta I, Kotoula V, Lakis S, Dimopoulos MA, Skarlos D, Pectasides D, Fountzilas G, Beckmann MW, Hein A, Ruebner M, Ekici AB, Hartmann A, Schulz-Wendtland R, Renner SP, Janni W, Rack B, Scholz C, Neugebauer J, Andergassen U, Lux MP, Haeberle L, Clarke C, Pathmanathan N, Rudolph A, Flesch-Janys D, Nickels S, Olson JE, Ingle JN, Olswold C, Slettedahl S, Eckel-Passow JE, Anderson SK, Visscher DW, Cafourek VL, Sicotte H, Prodduturi N, Weiderpass E, Bernstein L, Ziogas A, Ivanovich J, Giles GG, Baglietto L, Southey M, Kosma VM, Fischer HP; GENICA Network, Reed MW, Cross SS, Deming-Halverson S, Shrubsole M, Cai Q, Shu XO, Daly M, Weaver J, Ross E, Klemp J, Sharma P, Torres D, Rüdiger T, Wölfling H, Ulmer HU, Försti A, Khoury T, Kumar S, Pilarski R, Shapiro CL, Greco D, Heikkilä P, Aittomäki K, Blomqvist C, Irwanto A, Liu J, Pankratz VS, Wang X, Severi G, Mannermaa A, Easton D, Hall P, Brauch H, Cox A, Zheng W, Godwin AK, Hamann U, Ambrosone C, Toland AE, Nevanlinna H, Vachon CM, Couch FJ. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. *Carcinogenesis* 2014; 35(5): 1012-1019.
- [24] Tindall EA, Bornman MS, van Zyl S, Segone AM, Monare LR, Venter PA, Hayes VM. Addressing the contribution of previously described genetic and epidemiological risk

- factors associated with increased prostate cancer risk and aggressive disease within men from South Africa. *BMC Urology* 2013; 13: 74.
- [25] Butoescu V, Ambroise J, Stainier A, Dekairelle AF, Gala JL, Tombal B. Does genotyping of risk-associated single nucleotide polymorphisms improve patient selection for prostate biopsy when combined with a prostate cancer risk calculator? *Prostate* 2014; 74(4): 365-371.
- [26] He C, Chasman DI, Dreyfus J, Hwang SJ, Ruiter R, Sanna S, Buring JE, Fernández-Rhodes L, Franceschini N, Hankinson SE, Hofman A, Lunetta KL, Palmieri G, Porcu E, Rivadeneira F, Rose LM, Splansky GL, Stolk L, Uitterlinden AG, Chanock SJ, Crisponi L, Demerath EW, Murabito JM, Ridker PM, Stricker BH, Hunter DJ. Reproductive aging-associated common genetic variants and the risk of breast cancer. *Breast Cancer Research* 2012; 14(2): R54.
- [27] Zheng J, Liu F, Lin X, Wang X, Ding Q, Jiang H, Chen H, Lu D, Jin G, Hsing AW, Shao Q, Qi J, Ye Y, Wang Z, Gao X, Wang G, Chu LW, Ouyang J, Huang Y, Chen Y, Gao Y, Shi R, Wu Q, Wang M, Zhang Z, Hu Y, Sun J, Zheng SL, Gao X, Xu C, Mo Z, Sun Y, Xu J. Predictive performance of prostate cancer risk in Chinese men using 33 reported prostate cancer risk-associated SNPs. *Prostate* 2012; 72(5): 577-583.